

KS  
#2

Docket No. 1359.1023/HJS

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

In re Application of:

Shoji HAYAKAWA et al.

Group Art Unit:

Serial No.:

Examiner:

Filed: January 19, 2000

For: **SPEAKER VERIFICATIN APPARATUS AND METHOD**

jc675 U.S. PTO  
09/487467  
01/19/00

**SUBMISSION OF CERTIFIED COPY OF PRIOR  
FOREIGN APPLICATION IN ACCORDANCE WITH  
THE REQUIREMENTS OF 37 C.F.R. § 1.55**

Assistant Commissioner for Patents  
Washington, D.C. 20231

Sir:

In accordance with the provisions of 37 C.F.R. § 1.55, the applicant(s) submit(s)  
herewith a certified copy of the following foreign application(s):

Japanese Patent Application No. 11-94734  
Filed: April 1, 1999

It is respectfully requested that the applicant(s) be given the benefit of the foreign filing  
date, as evidenced by the certified papers attached hereto, in accordance with the requirements  
of 35 U.S.C. § 119.

Respectfully submitted,  
STAAS & HALSEY LLP

Date: January 19, 2000

By: \_\_\_\_\_

  
H. J. Staps  
Registration No. 22,010

700 Eleventh Street, N.W.  
Suite 500  
Washington, D.C. 20001  
Telephone: (202) 434-1500  
Facsimile: (202) 434-1501

日本国特許庁  
PATENT OFFICE  
JAPANESE GOVERNMENT

Jc675 U.S. PTO  
09/487467  
01/19/00

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日

Date of Application:

1999年 4月 1日

出願番号

Application Number:

平成11年特許願第094734号

出願人

Applicant(s):

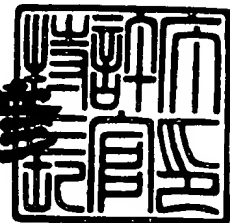
富士通株式会社

CERTIFIED COPY OF  
PRIORITY DOCUMENT

1999年11月26日

特許庁長官  
Commissioner,  
Patent Office

近藤隆彦



出証番号 出証特平11-3081866

【書類名】 特許願

【整理番号】 9990043

【提出日】 平成11年 4月 1日

【あて先】 特許庁長官 殿

【国際特許分類】 G10L 3/00

【発明の名称】 話者照合装置及び方法

【請求項の数】 6

【発明者】

    【住所又は居所】 神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内

    【氏名】 早川 昭二

【発明者】

    【住所又は居所】 愛知県名古屋市名東区高針台3-506

    【氏名】 板倉 文忠

【発明者】

    【住所又は居所】 愛知県名古屋市名東区八前2-816

    【氏名】 武田 一哉

【特許出願人】

    【識別番号】 000005223

    【氏名又は名称】 富士通株式会社

【代理人】

    【識別番号】 100095555

    【弁理士】

    【氏名又は名称】 池内 寛幸

    【電話番号】 06-6361-9334

【手数料の表示】

    【予納台帳番号】 012162

    【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9803089

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 話者照合装置及び方法

【特許請求の範囲】

【請求項 1】 個人 ID を入力する個人 ID 入力部と、

前記個人 ID 入力部で入力された前記個人 ID に対応する登録話者の音声を選択する話者選択部と、

話者の音声を格納しておく話者記憶部と、

話者の音声を入力する音声入力部と、

前記音声入力部で入力された音声を分析する音声分析部と、

前記音声分析部における分析結果及び前記話者記憶部に格納された前記音声に基づいて、前記入力された音声と前記登録話者の音声との照合距離、及び前記入力された音声と前記話者記憶部に格納されている前記登録話者を除いた他の登録話者の音声との話者間距離を計算する話者間距離計算部と、

前記入力された音声が入力された前記個人 ID に対応する登録話者と一致するか否かを判定する話者判定部を含む話者照合装置であって、

所定の詐称者受理誤り率をしきい値として入力する詐称者受理誤り率入力部と

前記話者間距離計算部において計算された前記話者間距離に基づいて前記話者間距離の確率分布を求める分布推定部をさらに含み、

前記話者判定部で、前記話者間距離の確率分布における入力された前記詐称者受理誤り率を示す領域内に、前記話者間距離計算部において計算された前記照合距離が含まれる場合に本人の音声であると判断することを特徴とした話者照合装置。

【請求項 2】 前記話者間距離の確率分布が正規分布関数であると仮定し、

前記話者判定部において、前記正規分布関数から求められる前記話者間距離の確率分布における詐称者受理誤り率を示す領域内に、前記話者間距離計算部において計算された照合距離が含まれる場合に本人の音声であると判断する請求項 1 記載の話者照合装置。

【請求項 3】 性別ごとに区分して前記話者間距離の確率分布を求めた請求項

1 記載の話者照合装置。

【請求項 4】 複数の正規分布の重み付け和として前記話者間距離の確率分布を求めた請求項 1 記載の話者照合装置。

【請求項 5】 個人 ID を入力する工程と、  
入力された前記個人 ID に対応する登録話者の音声を選択する工程と、  
話者の音声を入力する工程と、  
入力された音声を分析する工程と、  
分析結果及び音声に基づいて、前記入力された音声と前記登録話者の音声との照合距離、及び前記入力された音声と前記登録話者を除いた他の登録話者の音声との話者間距離を計算する工程と、

前記入力された音声が入力された前記個人 ID に対応する登録話者と一致するか否かを判定する工程を含む話者照合方法であって、

所定の詐称者受理誤り率をしきい値として入力する工程と、  
計算された前記話者間距離に基づいて前記話者間距離の確率分布を求める工程をさらに含み、

前記話者間距離の確率分布における入力された前記詐称者受理誤り率を示す領域内に、前記話者間距離計算部において計算された前記照合距離が含まれる場合に本人の音声であると判断することを特徴とする話者照合方法。

【請求項 6】 個人 ID を入力するステップと、  
入力された前記個人 ID に対応する登録話者の音声を選択するステップと、  
話者の音声を入力するステップと、  
入力された音声を分析するステップと、  
分析結果及び音声に基づいて、前記入力された音声と前記登録話者の音声との照合距離、及び前記入力された音声と前記登録話者を除いた他の登録話者の音声との話者間距離を計算するステップと、

前記入力された音声が入力された前記個人 ID に対応する登録話者と一致するか否かを判定するステップを含むコンピュータに実行させるプログラムを記録したコンピュータ読み取り可能な記録媒体であって、

所定の詐称者受理誤り率をしきい値として入力するステップと、

計算された前記話者間距離に基づいて前記話者間距離の確率分布を求めるステップをさらに含み、

前記話者間距離の確率分布における入力された前記詐称者受理誤り率を示す領域内に、前記話者間距離計算部において計算された前記照合距離が含まれる場合に本人の音声であると判断することを特徴としたコンピュータに実行させるプログラムを記録したコンピュータ読み取り可能な記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、事前に登録してある音声の特徴パラメータに基づいて、利用者本人か否かを音声によって判定する話者照合装置又は方法に関する。

【0002】

【従来の技術】

近年、コンピュータ技術の発展とともに、急速に通信環境についても整備されつつある。このような通信環境の整備に伴い、電話によるコンピュータアクセス (Computer Telephony Integration) が一般の家庭においても普通に行うことが可能になってきた。

【0003】

かかる電話によるコンピュータアクセス分野においては、プライバシーに関する情報や秘密保持義務を有する情報等に代表される、本人や特定の個人以外に知らせてはならない情報に対するアクセスを行う場合に問題がある。すなわち、例えばプッシュホンを用いる場合においては、パスワードを電話のボタン操作によって入力することで当該情報へのアクセス権限を取得することが可能であるが、パスワードを他人に知られてしまうと、本人でないにもかかわらず、当該情報に容易にアクセスできてしまうという問題である。そのため、本人に固有である音声を用いて、本人あるいは特定の個人であるか否かについて照合を行うことが求められている。かかるセキュリティ機能を確実に発揮させるためには、特に照合対象となる音声の登録や、入力された音声が本人の音声か否かを判別するしきい値の決定に際して、利用者にとって過度の負担とならないことが重要である。

## 【0004】

従来、本人の音声か否かを判別するしきい値には、固定した一定の値が用いられるのが主流であった。すなわち、図1に示すように、入力された音声と事前に登録されている音声との間で照合距離を計算し、あらかじめ設定したしきい値と比較して、当該しきい値よりも照合距離が同じか短い場合（図1の“-”）には本人であると、長い場合（図1の“+”）には他人であると判断するものである。

## 【0005】

かかるしきい値は、次のような値に設定することが望ましい。図2は、照合距離のしきい値を横軸として、本人ではないと棄却する判断が誤りであった場合の確率である本人棄却誤り率FR (False Rejection error rate) を縦軸にとる。一方、同じく照合距離を横軸として、詐称者であるとする判断が誤りであった場合の確率である詐称者受理誤り率FA (False Acceptance error rate) を縦軸にとる。しきい値を小さな値にすると、詐称者を誤って受理してしまう率FAは減るが、本人を誤って棄却してしまう率FRが高くなる。逆にしきい値を大きな値とすると、本人を誤って棄却してしまう率FRは小さくなるが、詐称者を誤って受理してしまう率FAは大きくなる。よって、かかる2つの誤り率の重要度に応じて、しきい値を適切な値に設定するのが望ましい。実験的には事後的にかかる2つの誤り率が等しくなる値をしきい値として評価するのが一般的である。

## 【0006】

## 【発明が解決しようとする課題】

しかしながら、上述した方法においては、しきい値を設定する際には本人棄却誤り率FRと詐称者受理誤り率FAの傾向について、あらかじめ知っておく必要が生じる。しかし、利用する前にかかる2つの誤り率の傾向を知ることは困難である。したがって、予備実験を行うことで近似値を模索したり、システムの利用時に逐次更新していく方法等が採用されている。予備実験を行う方法については、予備実験の実行時と実際の利用時における条件の差異によって、結局、利用時に再テストを行う必要がある場合が多い。また、本人棄却誤り率FRを算出するためには、本人（利用者）に何回も発声してもらう必要があり、これは利用者の

負担が大きく、実用的ではない。一方、システムの利用時に逐次更新していく方法についても、当該しきい値を逐次更新していくことは利用者にとっては負担が大きいという問題点があった。

【0007】

さらに、たとえ本人の音声であっても、時間が経過するにつれて変化することもあり、また雑音等のノイズが混入した場合にも正確な話者認識は一般に困難である。

【0008】

本発明は、上記問題点を解決すべく、実行環境の設定が使用者に過度の負担となることなく、かつ、精度良く話者を特定することのできる話者照合装置及び方法を提供することを目的とする。

【0009】

【課題を解決するための手段】

上記目的を達成するために本発明にかかる話者照合装置は、個人IDを入力する個人ID入力部と、個人ID入力部で入力された個人IDに対応する登録話者の音声を選択する話者選択部と、話者の音声を格納しておく話者記憶部と、話者の音声を入力する音声入力部と、音声入力部で入力された音声を分析する音声分析部と、音声分析部における分析結果及び話者記憶部に格納された音声に基づいて、入力された音声と登録話者の音声との照合距離、及び入力された音声と話者記憶部に格納されている登録話者を除いた他の登録話者の音声との話者間距離を計算する話者間距離計算部と、入力された音声が入力された個人IDに対応する登録話者と一致するか否かを判定する話者判定部を含む話者照合装置であって、所定の詐称者受理誤り率をしきい値として入力する詐称者受理誤り率入力部と、話者間距離計算部において計算された話者間距離に基づいて話者間距離の確率分布を求める分布推定部をさらに含み、話者判定部で、話者間距離の確率分布における入力された詐称者受理誤り率を示す領域内に、話者間距離計算部において計算された照合距離が含まれる場合に本人の音声であると判断することを特徴とする。

【0010】

かかる構成により、固定された照合距離のしきい値を用いるのではなく、システムの利用ごとに話者間距離の確率分布を計算し、その分布における詐称者所理誤り率に基づいてしきい値を決定することにより、統計的な確率分布理論値に近い話者判別基準を得ることができ、また音声入力環境が変動し、雑音が混入した場合であっても詐称者受理誤り率を理論値近傍に維持することができ、入力する音声の経年変化に影響されず、話者照合における照合精度を高いレベルで維持することができる。なお、本発明は、入力される音声の環境の相異や経年変化等の外的要因によって話者間距離自体は常に変動するのに対して、入力音声と他の登録話者との話者間距離との関係においては、その関係自体はあまり変化しないという経験的事実に基づくものである。

## 【 0 0 1 1 】

また、本発明は、上記のような話者照合装置の機能をコンピュータの処理ステップとして実行するソフトウェアを特徴とするものであり、具体的には、個人IDを入力する工程と、入力された個人IDに対応する登録話者の音声を選択する工程と、話者の音声を入力する工程と、入力された音声を分析する工程と、分析結果及び音声に基づいて、入力された音声と登録話者の音声との照合距離、及び入力された音声と登録話者を除いた他の登録話者の音声との話者間距離を計算する工程と、入力された音声が入力された個人IDに対応する登録話者と一致するか否かを判定する工程を含む話者照合方法であって、所定の詐称者受理誤り率をしきい値として入力する工程と、計算された話者間距離に基づいて話者間距離の確率分布を求める工程をさらに含み、話者間距離の確率分布における入力された詐称者受理誤り率を示す領域内に、話者間距離計算部において計算された照合距離が含まれる場合に本人の音声であると判断する話者照合方法並びにそのような工程をプログラムとして記録したコンピュータ読み取り可能な記録媒体であることを特徴とする。

## 【 0 0 1 2 】

かかる構成により、コンピュータ上へ当該プログラムをロードさせ実行することで、固定された照合距離のしきい値を用いるのではなく、システムの利用ごとに話者間距離の確率分布を計算し、その分布における詐称者所理誤り率に基づい

てしきい値を決定することにより、統計的な確率分布理論値に近い話者判別基準を得ることができ、また音声入力環境が変動し、雑音が混入した場合であっても詐称者受理誤り率を理論値近傍に維持することができ、入力する音声の経年変化に影響されず、話者照合における照合精度を高いレベルで維持することができる話者照合装置を実現することが可能となる。

【0013】

【発明の実施の形態】

以下、本発明の実施の形態にかかる話者照合装置について、図面を参照しながら説明する。図3は本発明の実施の形態にかかる話者照合装置の原理構成図である。図3において、31は個人ID入力部を、32は話者テンプレート選択部を、33は話者テンプレート記憶部を、34は音声入力部を、35は音声分析部を、36Aは照合距離計算部を、36Bは話者間距離計算部を、37は分布推定部を、38は詐称者受理誤り率入力部を、39は話者判定部を、それぞれ示す。

【0014】

図3において、システム使用時に個人IDをID入力部31から入力し、話者テンプレート選択部32において事前に話者テンプレート記憶部33に登録してある複数の話者についてのテンプレートから個人IDに対応するテンプレートを選択し、照合距離計算部36Aに送信する。同時に、個人IDに対応する話者を除いた登録話者のテンプレートについて話者間距離計算部36Bへ送信する。

【0015】

次に、音声分析部35では、音声入力部34で入力された音声を、話者照合のための特徴パラメータに変換した上で、照合距離計算部36A及び話者間距離計算部36Bへ送信する。そして、照合距離計算部36Aでは個人IDに対応する話者の音声テンプレートと入力音声の特徴パラメータとの距離 $d_{id}$ を計算する。

【0016】

一方、話者間距離計算部36Bでは他の登録話者N人の音声テンプレートと入力音声の特徴パラメータとの距離 $d_1$ 、 $d_2$ 、…、 $d_N$ を計算して、分布推定部37へ渡す。分布推定部37では、受け取ったN個の他の登録話者との間の距離計算結果 $d_1$ 、 $d_2$ 、…、 $d_N$ を用いて、入力された個人IDと対応する登録話者以

外の登録話者の音声と入力音声との話者間距離に関する確率分布関数  $F(d)$  を推定し、話者判定部 39 へ渡す。

【0017】

確率分布関数  $F(d)$  が推定できれば、確率密度関数  $f(d)$  を求めることができ、確率密度関数  $f(d)$  で示される関数の面積が確率値を示すことになる。なお、確率分布関数  $F(d)$  と確率密度関数  $f(d)$  の間には(数1)のような関係がある。

【0018】

【数1】

$$F(x) = \int_{-\infty}^x f(t) dt$$

【0019】

したがって話者判定部 39 においては、確率密度関数  $f(d)$  に基づいて、詐称者受理誤り率入力部 38 で事前に指定された詐称者を本人とみなす危険率  $p$  を示す領域内に個人 ID に対応する話者との話者間距離  $d_{id}$  が含まれる場合に本人であるとみなし、含まれない場合には詐称者であると判断する。これを確率分布関数  $F(d)$  に基づいて判断すると、 $F(d_{id}) < p$  ならば本人であると判断し、 $F(d_{id}) \geq p$  ならば詐称者であると判断する。

【0020】

話者判定部 39 における話者判定方法のイメージ図を図 4 に示す。図 4 のように確率密度関数  $f(d)$  が求まっている場合に、詐称者を本人とみなす危険率  $p$  を示す領域とは図 4 の斜線部で表される領域となる。すなわち、詐称者を本人とみなす危険率  $p$  を定めることで、詐称者を本人とみなす危険率が指定された危険率  $p$  より小さい範囲内に照合距離  $d_{id}$  が存在する場合に本人と判断するものである。

【0021】

次に、図 5 は本発明の一実施例にかかる話者照合装置における話者照合時のブロック構成図である。図 5 において、51A 及び 51B は DP マッチング部を、52 は統計量計算部を、53 は話者判定部を、54 は詐称者受理誤り率入力部を、それぞれ示す。

#### 【0022】

図 5 においても図 3 と同様に、システム使用時に個人 ID を ID 入力部 31 から入力し、話者テンプレート選択部 32 において事前に話者テンプレート記憶部 33 に登録してある複数の話者についてのテンプレートから個人 ID に対応するテンプレートを選択し、DP マッチング部 51A に送信する。同時に、個人 ID に対応する話者を除いた登録話者のテンプレートについて DP マッチング部 51B へ送信する。ここで、DP とは動的計画法 (Dynamic Programming) をいう。

#### 【0023】

次に、音声分析部 35 では、音声入力部 34 で入力された音声を、話者照合のための特徴パラメータに変換した上で、DP マッチング部 51A 及び DP マッチング部 51B へ送信する。そして、DP マッチング部 51A では個人 ID に対応する話者の音声テンプレートと入力音声の特徴パラメータとの距離  $d_{id}$  を計算する。

#### 【0024】

一方、DP マッチング部 51B では N 人の他の登録話者に関する音声テンプレートと入力音声の特徴パラメータとの距離  $d_1, d_2, \dots, d_N$  を計算して、統計量計算部 52 へ渡す。統計量計算部 52 では、受け取った N 個の他の登録話者に関する距離計算結果  $d_1, d_2, \dots, d_N$  を用いて、話者間距離の平均値  $\mu$  と標準偏差  $\sigma$  を推定して、話者判定部 53 へ渡す。話者判定部 53 では、他の登録話者との距離についての平均値  $\mu$  と標準偏差  $\sigma$  を用いて正規分布を構成する。

#### 【0025】

正規分布であれば、その平均値  $\mu$  から  $\alpha \cdot \sigma$  だけ離れた点における確率分布関数  $F(d)$  を  $\alpha$  によって定めることができる。したがって、入力音声との照合距離  $d_{id}$  が事前に指定する詐称者を本人とみなす危険率  $p$  を示す領域内に含まれるか否かは、 $(\mu - \alpha \cdot \sigma)$  よりも小さな領域に照合距離  $d_{id}$  が存在するか否かを

検討することで本人か否かを判別することができる。具体的には、 $(\mu - \alpha \cdot \sigma)$  と  $d_{id}$  を比較して、 $(\mu - \alpha \cdot \sigma)$  と  $d_{id}$  が同じであるか、 $(\mu - \alpha \cdot \sigma)$  よりも  $d_{id}$  が小さい場合に本人であると判断し、 $(\mu - \alpha \cdot \sigma)$  よりも  $d_{id}$  が大きい場合には詐称者であると判断する。なお、確率分布を正規分布と仮定する場合には、詐称者受理誤り率入力部 54 では、事前に詐称者を本人とみなす危険率  $p$  に対応する  $\alpha$  を入力することになる。

## 【0026】

本実施の形態においては、事前にテンプレート形式で特徴パラメータを登録しておき、DP マッチングにより求めた話者間距離に基づいて他の登録話者との確率分布を推定しているが、特にこの方法に限定されるものではなく、例えば Hidden Markov Model 等の確率モデルから出力される確率値に基づいて確率分布を推定する方法等でも良い。

## 【0027】

また、話者テンプレート記憶部 33 内において、事前に性別ごとに区別して話者を分割しておき、個人 ID が示す話者が男性の場合には他の男性話者に関する話者テンプレートを、女性ならば他の女性話者に関する話者テンプレートをを用いて確率分布を推定することで、正規分布関数表から求まる誤り率とより良く合致させることが可能となる。

## 【0028】

さらに、本実施の形態においては、単一の正規分布として話者間距離の確率分布を推定しているが、複数の正規分布に関するの重み付け和で構成される混合正規分布やその他の一般的な確率分布として推定する方法でも良い。

## 【0029】

次に、本実施の形態による効果を以下の実験結果に基づいて確認する。まず、図 6 は男性話者 15 人について本実施の形態にかかる話者照合方法を用いて照合した結果を示す図である。

## 【0030】

図 6 において、横軸は事前に指定する詐称者誤り率にしたがって正規分布関数から求まる  $\alpha$  である。また、詐称者誤り率の理論値を実線で示しているが、これ

は話者間距離の分布を正規分布であると仮定しているので、正規分布関数  $\Phi(\alpha)$  を用いて  $1 - \Phi(\alpha)$  として計算できる。

【0031】

さらに、FA (False Acceptance) は誤って詐称者を受理してしまう確率である詐称者受理誤り率を、FR (False Rejection) は誤って本人を棄却してしまう確率である本人棄却誤り率を示している。

【0032】

図6では、実線が詐称者受理誤り率の理論値を示し、短破線が実験によって求めたFRを、長破線が実験によって求めたFAを、それぞれ示している。図6に示すように、実線と長破線とは、ほとんど一致していることから、詐称者受理誤り率については理論値と大きく相異なることはなく、詐称者受理誤り率に基づいて話者照合を行うという本実施の形態にかかる話者照合方法の照合精度は高いことが期待できる。

【0033】

また、図7においては、照合すべき音声にSNR (Signal Noise Ratio) が約20dBの白色雑音を加えた場合の照合結果を図6と同様に示している。ここでSNRが約20dBとは、信号10に対して雑音1の割合で雑音が混入する程度を示す。また、実線が詐称者受理誤り率の理論値を示し、FR (noisy) が白色雑音混入時のFRを、FR (clean) が白色雑音のない時のFRを、FA (noisy) が白色雑音混入時のFAを、FA (clean) が白色雑音のない時のFAを、それぞれ示す。

【0034】

図7の実験結果によると、FRについては白色雑音の有無によって本人を棄却してしまう確率である本人棄却誤り率が大きく変動してしまうことがわかる。反面、FAについては、白色雑音が有ろうと無かろうと詐称者受理誤り率の理論値である実線とFAに関する実験結果とがほとんど一致していることに変わりはない。したがって、音声入力環境が相異なることから予期せぬ雑音が音声と同時に入力されたとしても、詐称者受理誤り率に基づいて話者照合を行うという本実施の形態にかかる話者照合方法の照合精度は高いレベルで保持されることが期待で

きる。

【0035】

さらに、図8においては、入力音声と登録音声の入力時期差を3ヶ月ごとに増やしていった場合の照合結果を図6と同様に示している。図8において、実線が詐称者受理誤り率の理論値を示し、FR、FAそれぞれについて、3ヶ月、6ヶ月、9ヶ月、12ヶ月経過時のFR、FA実験結果を、それぞれ示す。

【0036】

図8の実験結果によると、FRについては入力時期によって本人を棄却してしまう確率である本人棄却誤り率が大きく変動してしまうことがわかる。反面、FAについては、入力時期が異なろうと詐称者受理誤り率の理論値である実線と3ヶ月経過ごとのFAを示す破線とがほとんど一致していることに変わりはない。

【0037】

したがって、話者の入力時期が異なることで音声の特徴が変化しようと、他の登録話者との話者間距離自体については大きな変化が生じず、詐称者受理誤り率に基づいて話者照合を行うという本実施の形態にかかる話者照合方法の照合精度は高いレベルで維持されるとともに、一度登録した話者テンプレートについては利用機会ごとに更新する必然性もなく、利用者にとって過度の負担がかかることもない。

【0038】

次に、図9は本発明の一実施例にかかる話者照合装置における話者登録時のブロック構成図である。図9において、91は登録個人ID入力部を、92は登録音声入力部を、93は登録音声分析部を、94は背景雑音入力部を、95は雑音付加部を、96は他の登録話者に関する音声データベースを、それぞれ示す。

【0039】

図9において、登録個人ID入力部91から登録話者の個人IDを入力して、当該登録話者の音声を登録音声入力部92で入力する。登録音声入力部92で入力された音声は、登録音声分析部93で特徴パラメータに変換され、登録話者の音声テンプレートとして、個人ID情報と連携して話者テンプレート記憶部33に記憶される。

## 【0040】

次に、他の登録話者の音声データベース入力環境と当該登録話者の入力環境を一致させるために背景雑音入力部 94 で背景雑音を入力して、事前に登録している他の登録話者に関する音声データベース 96 内の音声データに入力した背景雑音を雑音付加部 95 で付加する。ここで、背景雑音とは音声を入力する際に同時に入力されてしまう雑音を意味する。実際の入力は、音声の入力前後での音声が入っていない状態における雑音のみを入力する。そして、登録音声分析部 93 において、個人 ID に対応した入力音声と同様に特徴パラメータに変換して、他の登録話者の音声テンプレートとして登録話者の音声テンプレートと同時に記憶する話者テンプレート記憶部 33 に記憶する。

## 【0041】

かかる構成とすることにより、他の登録話者の音声入力環境が登録話者の音声入力環境と大きく相異なることを未然に防止することができ、例えば詐称者の音声入力環境が他の登録話者の音声入力環境よりも登録話者の音声入力環境に近似している場合であっても、誤って詐称者を本人であると判断することを回避することができる。

## 【0042】

この場合、入力環境の調整は、音声データ自体について行うことに限定されるものではなく、音声の特徴パラメータに変換してから行っても良い。また、他の登録話者が Hidden Markov Model 等の確率モデルで表現されている場合においては、確率的な掛け合わせによって環境調整を実施しても良い。

## 【0043】

以上のように本実施の形態によれば、詐称者受理誤り率に基づいて話者照合を行うことで、統計的な確率分布理論値に近い話者判断基準を得ることができ、また音声入力環境が変動し、雑音が入力した場合であっても詐称者受理誤り率を理論値近傍に維持することができ、入力する音声の経年変化に影響されず、話者照合における照合精度を高いレベルで維持することができる。

## 【0044】

次に、本発明の実施の形態にかかる話者照合装置を実現するプログラムの処理

の流れについて説明する。図 10 及び図 11 に本発明の実施の形態にかかる話者照合装置を実現するプログラムの処理の流れ図を示す。

【0045】

まず、図 10 は本発明の実施の形態にかかる話者照合装置における話者照合時の処理流れ図である。図 10 において、個人 ID、音声、及び詐称者受理誤り率を利用者が入力する（ステップ S101）。詐称者受理誤り率については、システム管理者が、事前に入力しておき設定値としておくのが一般的である。

【0046】

そして、個人 ID に基づいて、個人 ID に対応した登録話者と、その他の登録話者との選別を行う（ステップ S102）。個人 ID に対応した登録話者については入力音声との照合距離を求めるのに使用し、その他の登録話者については話者間距離の確率分布を求めるのに使用するためである。

【0047】

そこで、入力された音声の特徴パラメータを抽出し（ステップ S103）、個人 ID に対応した登録話者との照合距離及びその他の登録話者との話者間距離を計算する（ステップ S104）。その他の登録話者との話者間距離の計算結果を用いることで、話者間距離の確率分布を推定することができる（ステップ S105）。

【0048】

話者間距離の確率分布が求まると、かかる確率分布上において、詐称者受理誤り率を示す領域を求めることができ、当該領域内に求まった個人 ID に対応した登録話者との照合距離が含まれているか否かを判断する（ステップ S106）。当該領域内に求まった個人 ID に対応した登録話者との照合距離が含まれている場合には入力された音声が個人 ID により特定される本人であると判断され（ステップ S107）、当該領域内に求まった個人 ID に対応した登録話者との照合距離が含まれていない場合には詐称者であると判断できる（ステップ S108）。

【0049】

次に、図 11 は本発明の実施の形態にかかる話者照合装置における話者登録時

の処理流れ図である。図 11 において、個人 ID、音声、及び背景雑音データを利用者が入力する（ステップ S111）。

【0050】

そして、他の登録話者の音声データを入手する（ステップ S112）。入手方法については特に限定されるものではないが、事前に他の登録話者に関する音声データのデータベースを構築しておく方が良い。

【0051】

次に入手した他の登録話者の音声データに入力された背景雑音を付加する（ステップ S113）。こうすることで、音声入力環境の相異を最小限にとどめることが可能となる。

【0052】

そこで、入力された音声及び雑音の付加された他の登録話者の音声データについて、各々の特徴パラメータを抽出する（ステップ S114）。個人 ID に対応した入力音声の特徴パラメータは話者音声テンプレートとして登録され、他の登録話者についての音声の特徴パラメータは他の登録話者との話者間距離を計算するのに用いる音声テンプレートとして同時に登録される（ステップ S115）。

【0053】

なお、本発明の実施の形態にかかる話者照合装置を実現するプログラムを記憶した記録媒体は、図 12 に示す記録媒体の例に示すように、CD-ROM やフロッピーディスク等の可搬型記録媒体だけでなく、通信回線の先に備えられた他の記憶装置や、コンピュータのハードディスクや RAM 等の記録媒体のいずれでも良く、プログラム実行時には、プログラムはローディングされ、主メモリ上で実行される。

【0054】

また、本発明の実施の形態にかかる話者照合装置により生成された話者テンプレート等を記録した記録媒体も、図 12 に示す記録媒体の例に示すように、CD-ROM やフロッピーディスク等の可搬型記録媒体だけでなく、通信回線の先に備えられた他の記憶装置や、コンピュータのハードディスクや RAM 等の記録媒体のいずれでも良く、例えば本発明にかかる話者照合装置を利用する際にコンピ

ュータにより読み取られる。

【 0 0 5 5 】

【発明の効果】

以上のように本発明にかかる話者照合装置によれば、詐称者受理誤り率に基づいて話者照合を行うことで、統計的な確率分布理論値に近い話者判断基準を得ることができ、また音声入力環境が変動し、雑音が混入した場合であっても詐称者受理誤り率を理論値近傍に維持することができ、入力する音声の経年変化に影響されず、話者照合における照合精度を高いレベルで維持することができる。

【図面の簡単な説明】

【図 1】 話者照合の一般概念図

【図 2】 従来の話者照合方法におけるしきい値特定方法の説明図

【図 3】 本発明の実施の形態にかかる話者照合装置の構成ブロック図

【図 4】 本発明の実施の形態にかかる話者照合装置におけるしきい値特定方法の説明図

【図 5】 本発明の一実施例にかかる話者照合装置の話者照合時における構成ブロック図

【図 6】 本発明の一実施例にかかる話者照合方法を用いた実験結果を示す図

【図 7】 本発明の一実施例にかかる話者照合方法を用いた実験結果を示す図

【図 8】 本発明の一実施例にかかる話者照合方法を用いた実験結果を示す図

【図 9】 本発明の一実施例にかかる話者照合装置における話者登録時の構成ブロック図

【図 1 0】 本発明の実施の形態にかかる話者照合装置における話者照合時の処理流れ図

【図 1 1】 本発明の実施の形態にかかる話者照合装置における話者登録時の処理流れ図

【図 1 2】 記録媒体の例示図

【符号の説明】

3 1 個人 I D 入力部

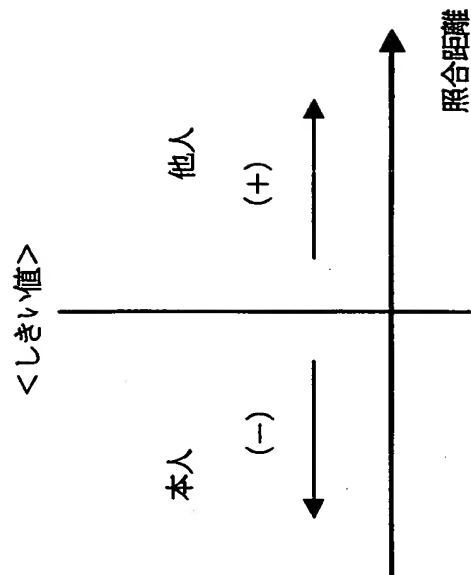
3 2 話者テンプレート選択部

- 3 3 話者テンプレート記憶部
- 3 4 音声入力部
- 3 5 音声分析部
- 3 6 A 照合距離計算部
- 3 6 B 話者間距離計算部
- 3 7 分析推定部
- 3 8、5 4 詐称者受理誤り率入力部
- 3 9、5 3 話者判定部
- 5 1 A、5 1 B DPマッチング部
- 5 2 統計量計算部
- 9 1 登録個人ID入力部
- 9 2 登録音声入力部
- 9 3 登録音声分析部
- 9 4 背景雑音入力部
- 9 5 雑音付加部
- 9 6 他の登録話者に関する音声データベース
- 1 2 1 回線先の記憶装置
- 1 2 2 CD-ROMやフロッピーディスク等の可搬型記録媒体
  - 1 2 2-1 CD-ROM
  - 1 2 2-2 フロッピーディスク
- 1 2 3 コンピュータ
- 1 2 4 コンピュータ上のRAM/ハードディスク等の記録媒体

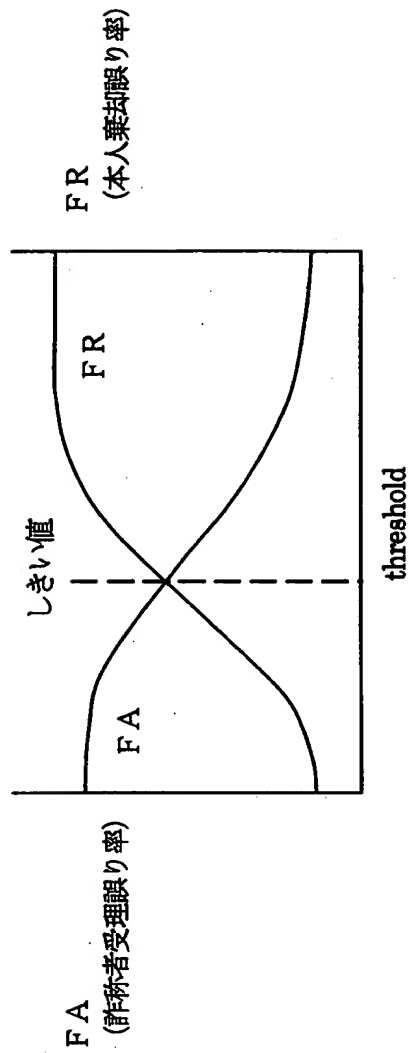
【書類名】

図面

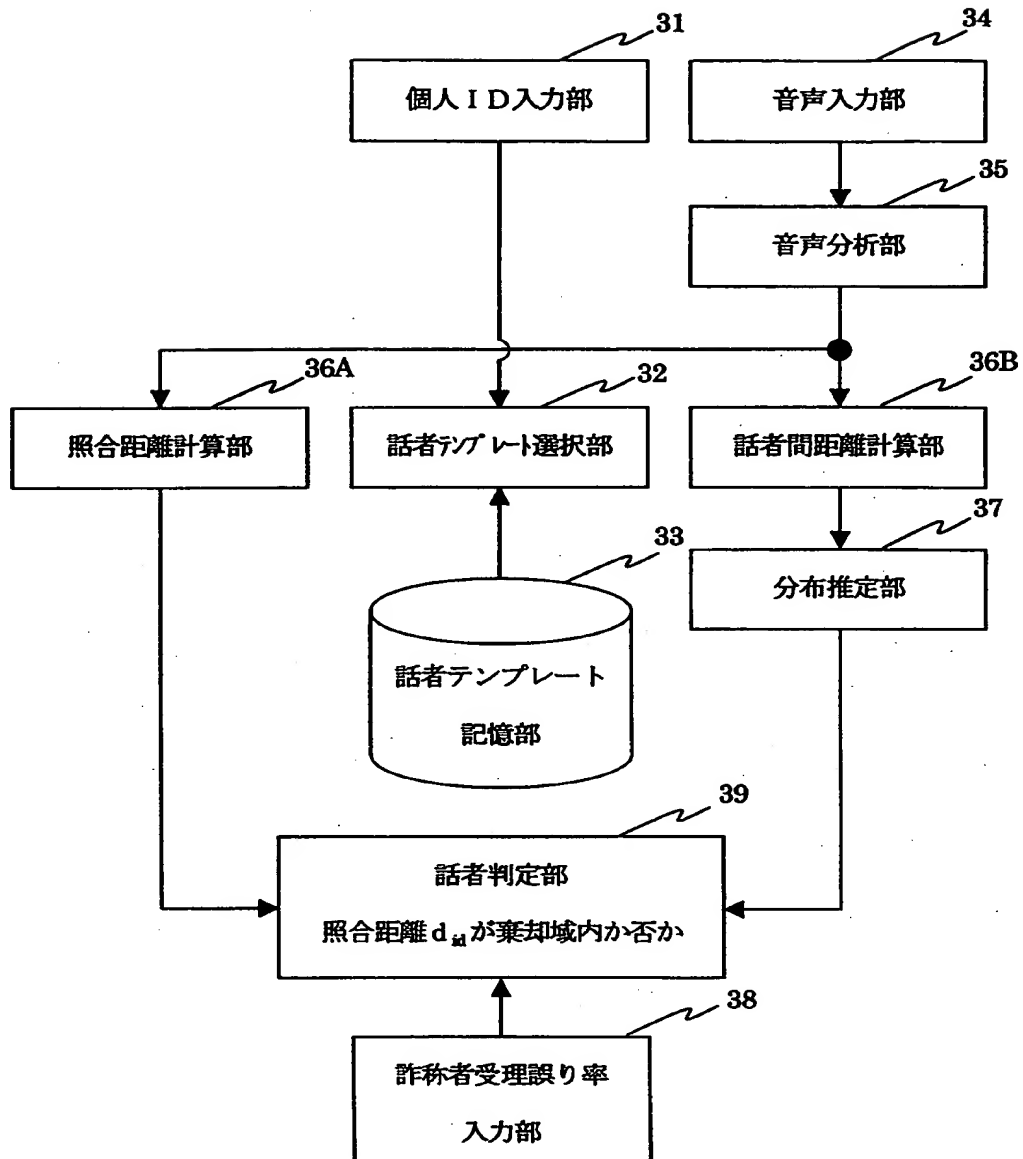
【図 1】



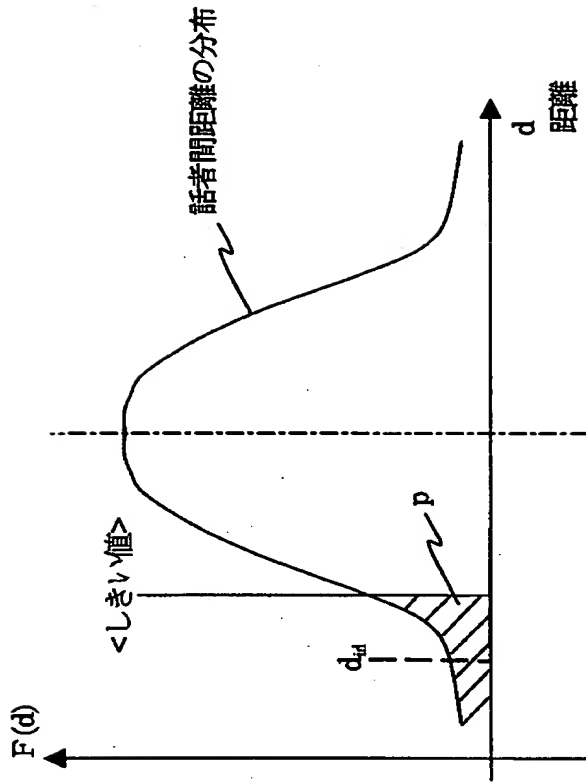
【図 2】



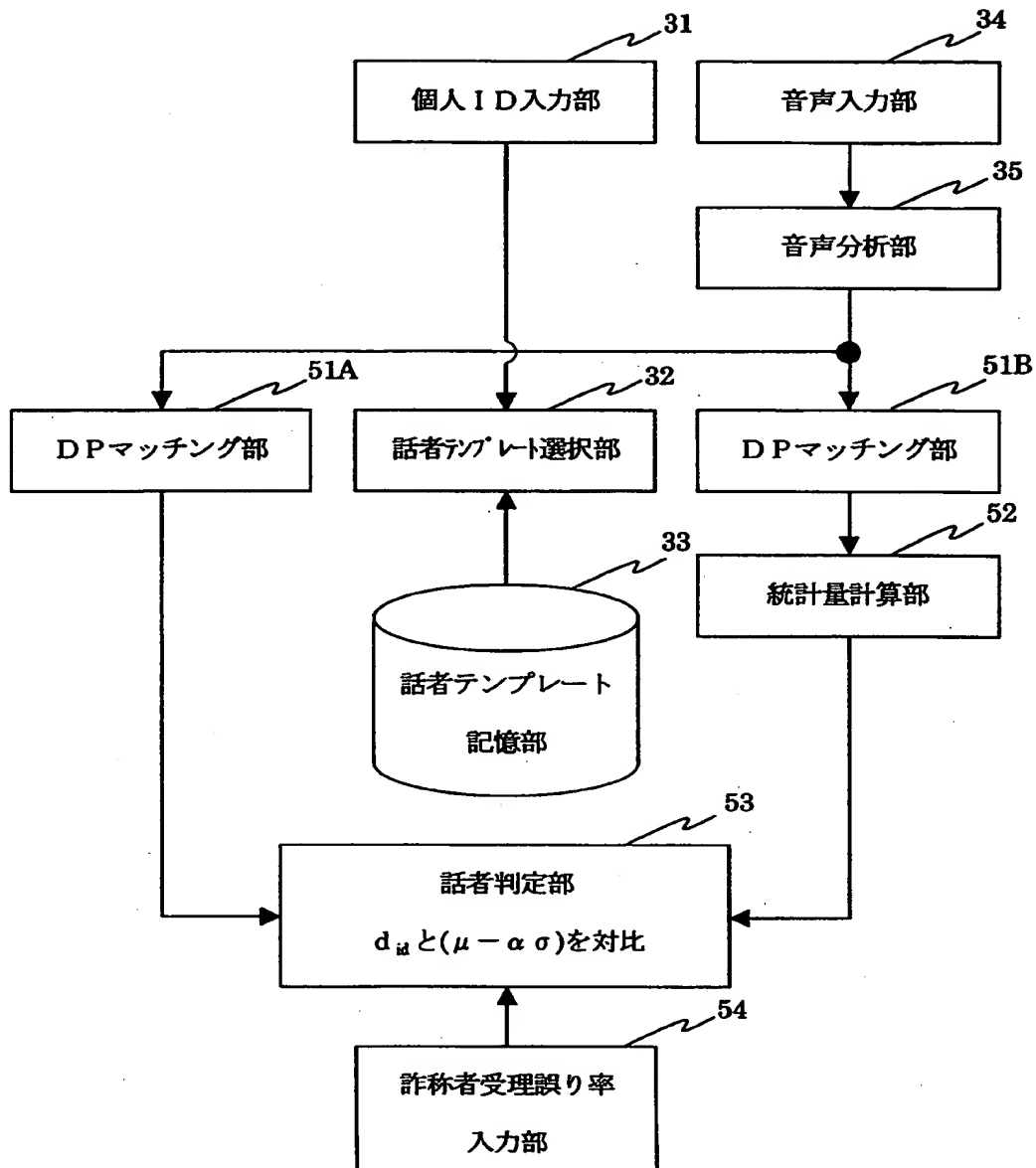
【図 3】



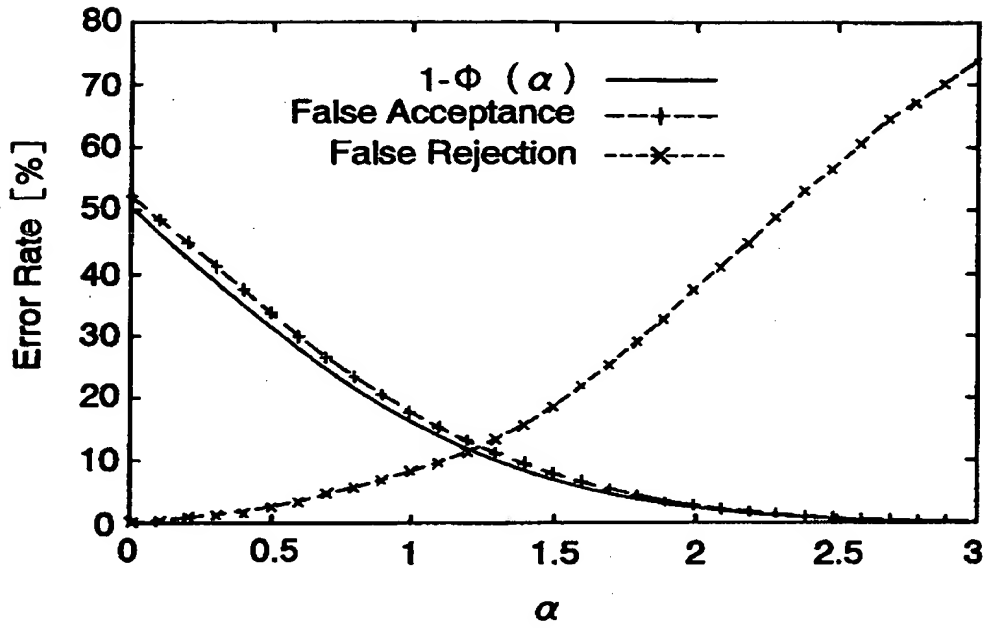
【図 4】



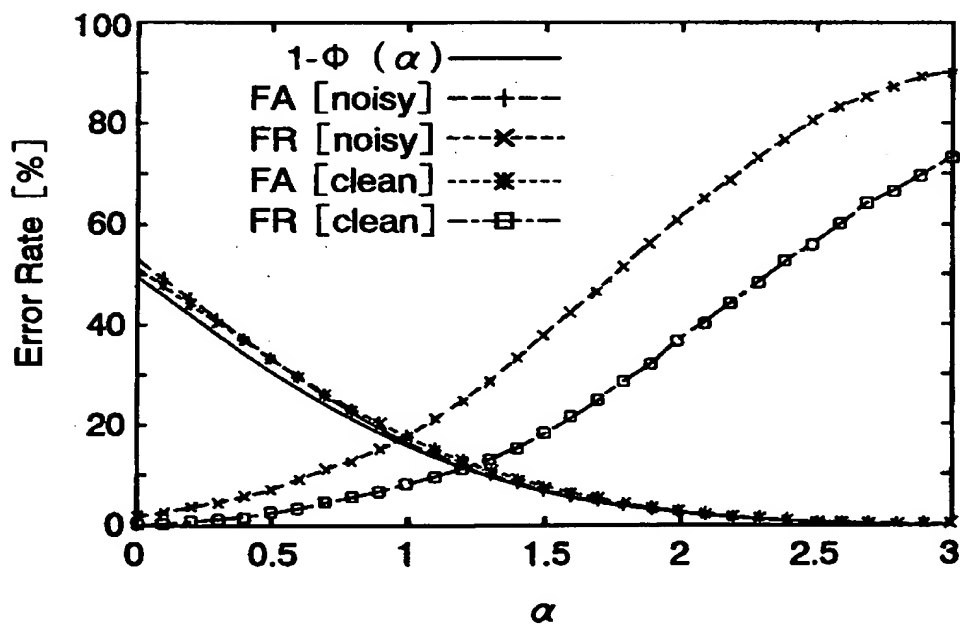
【図 5】



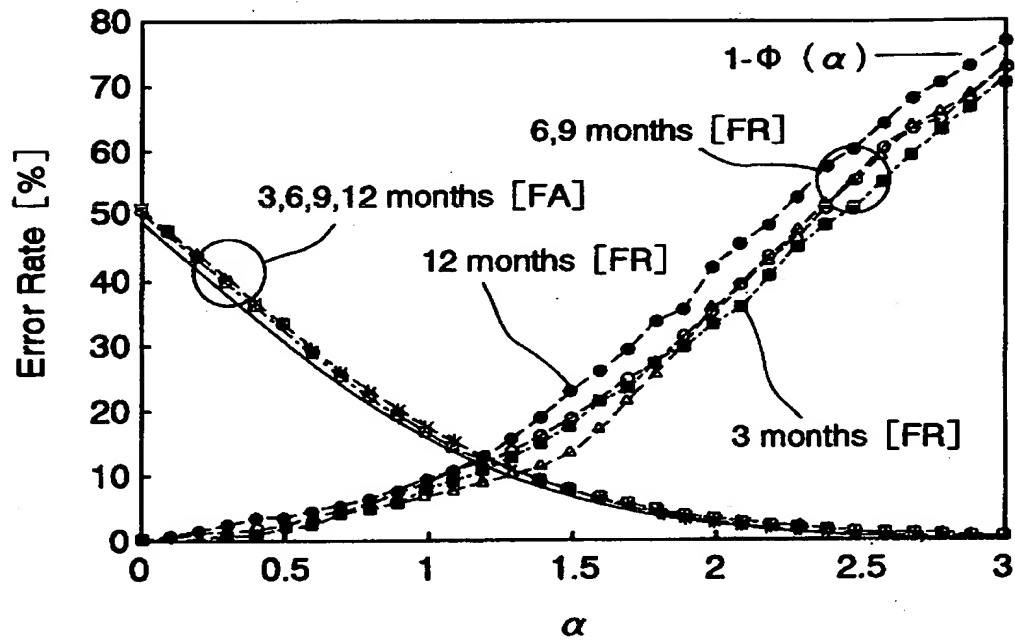
【図 6】



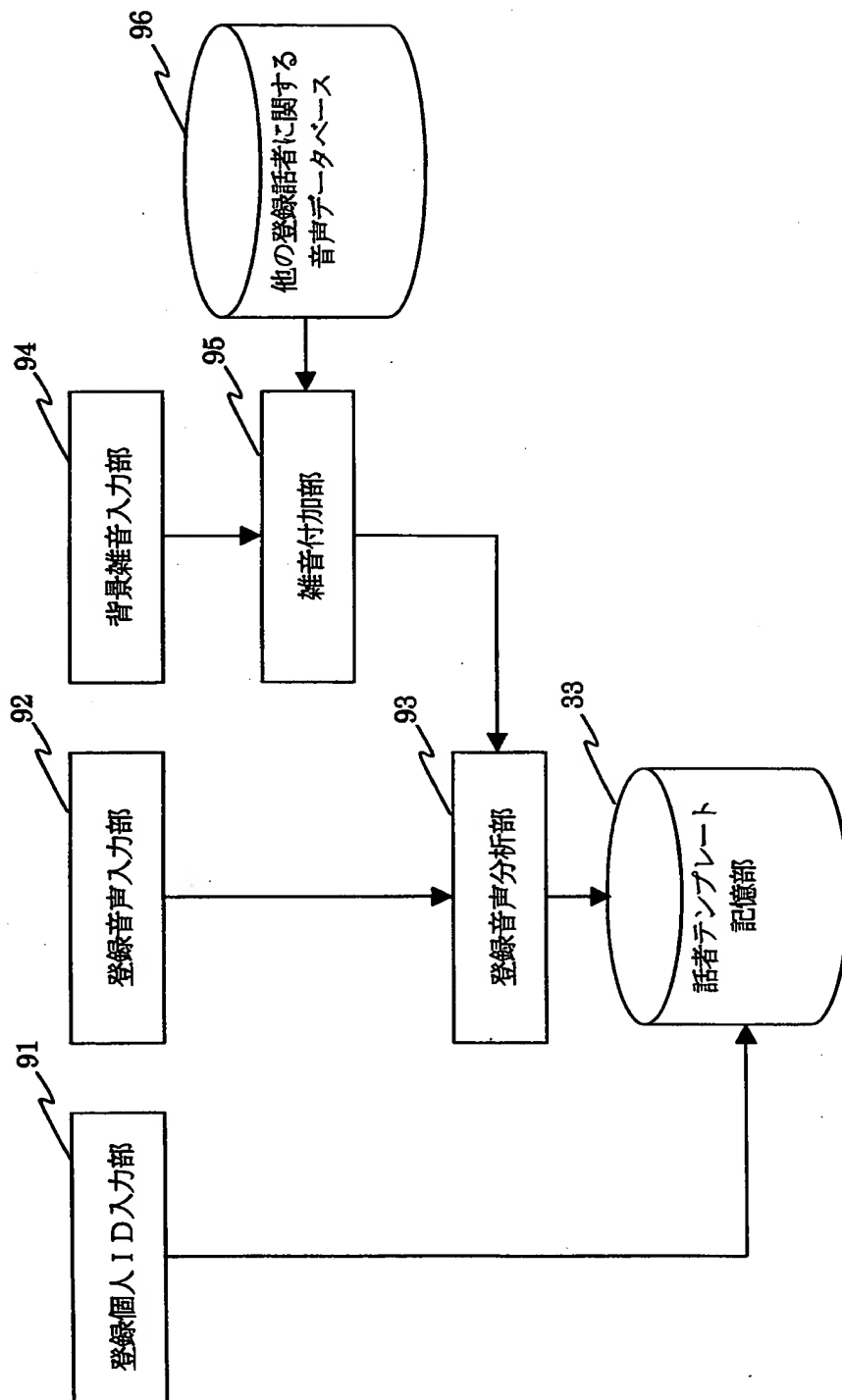
【図 7】



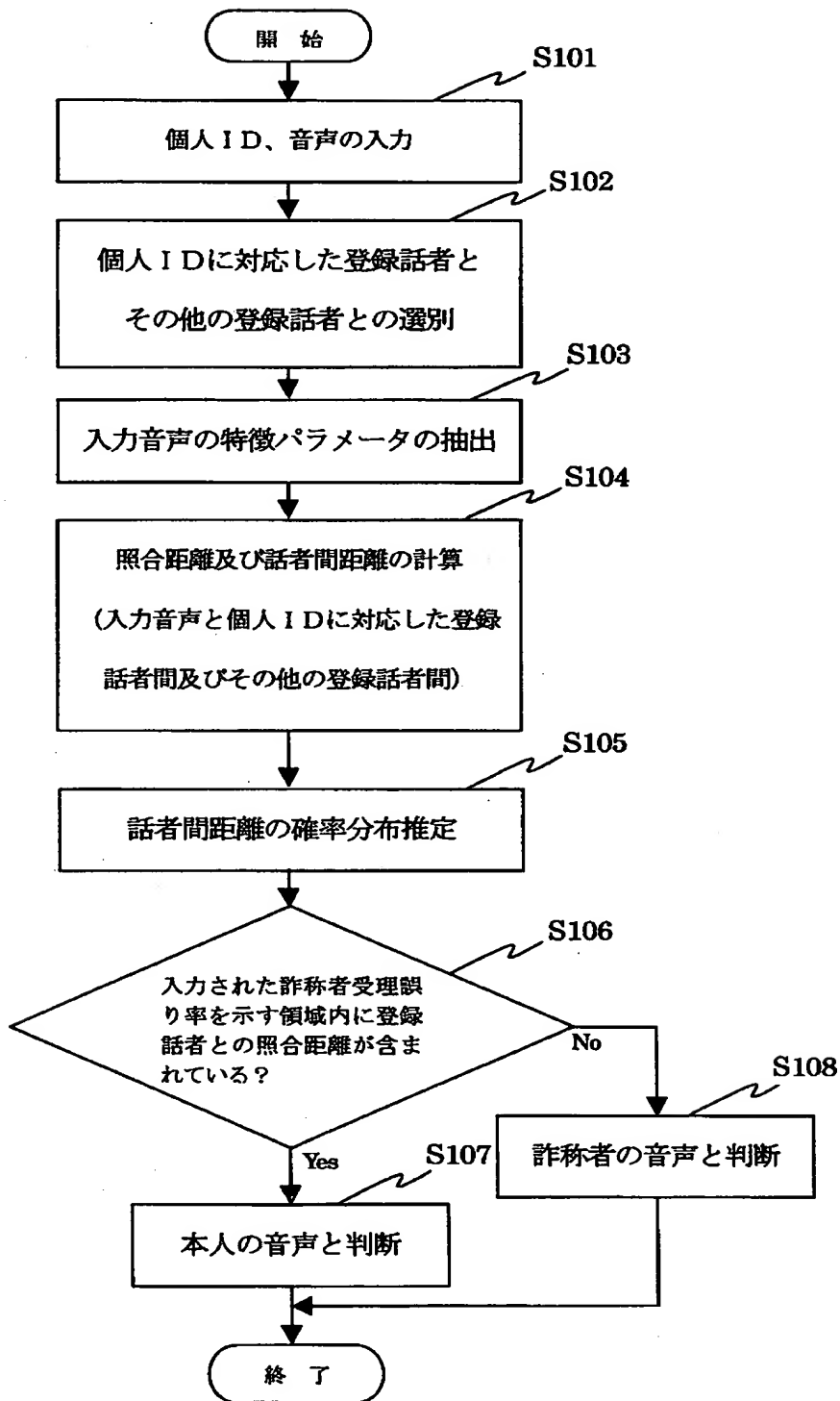
【図 8】



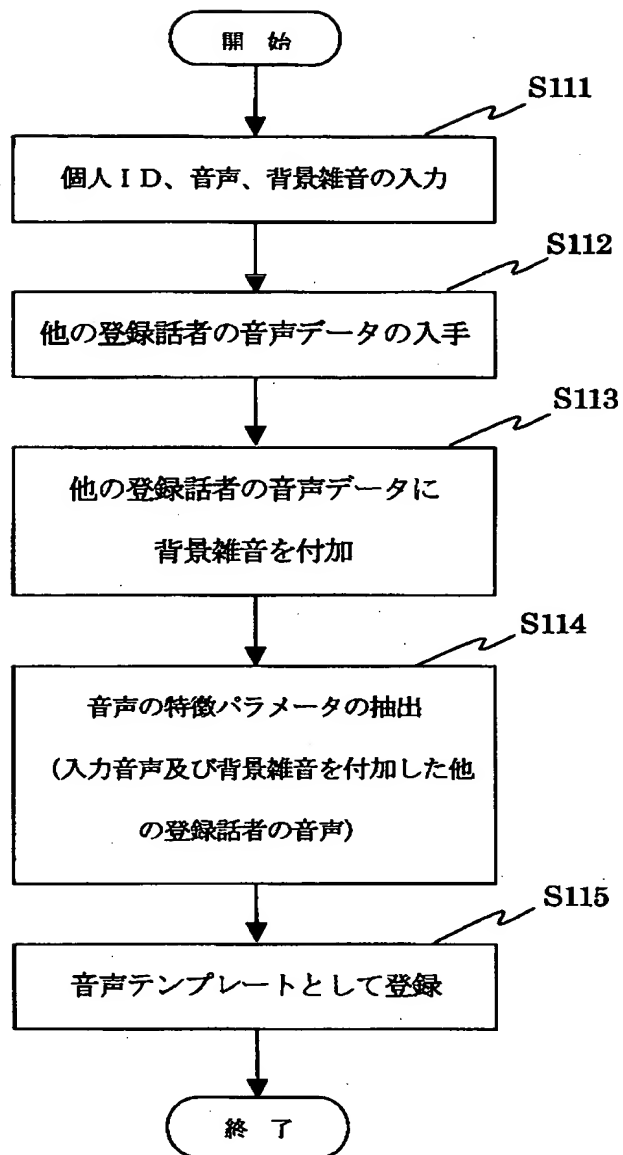
【図9】



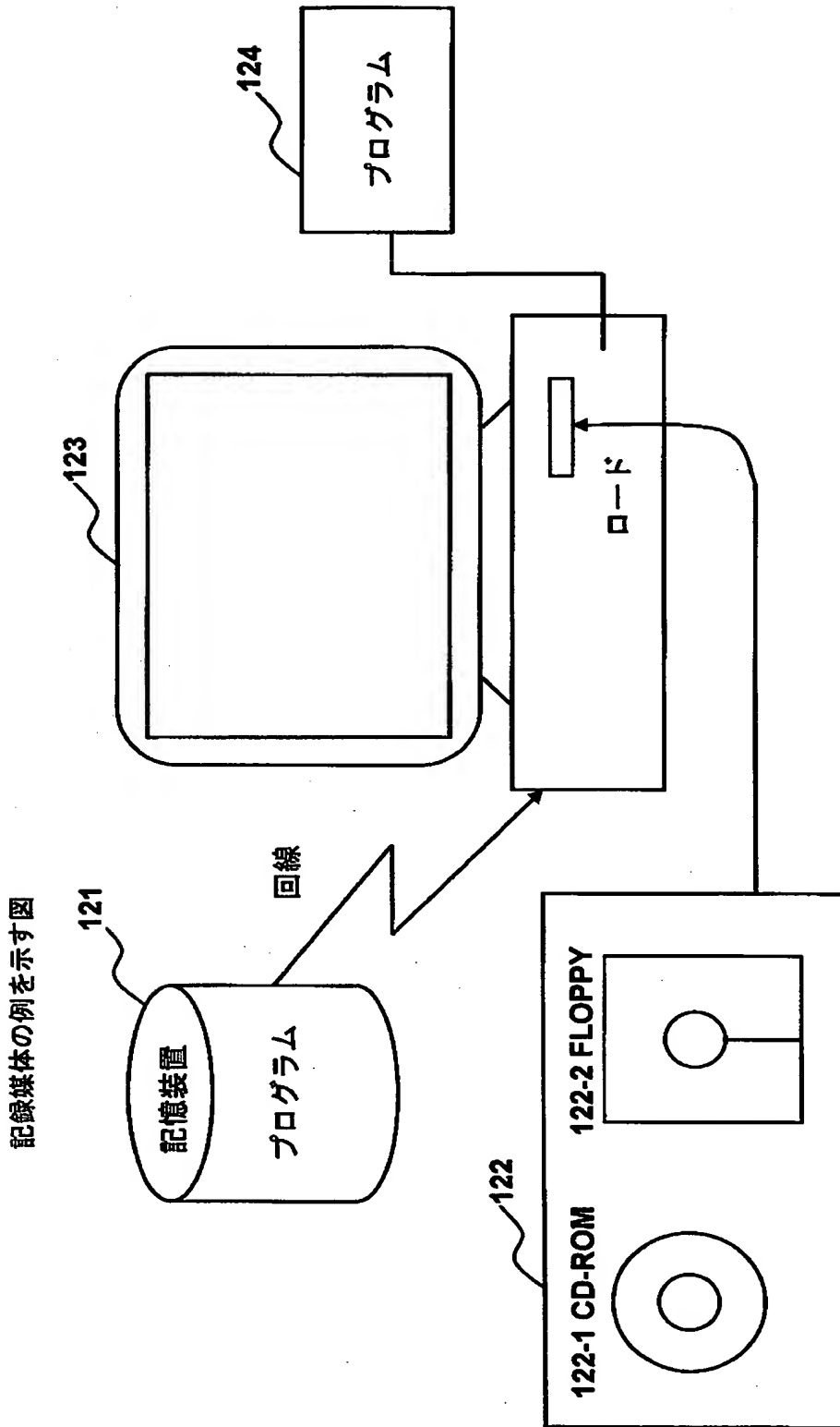
【図 10】



【図 1 1】



【図 1 2】



【書類名】 要約書

【要約】

【課題】 実行環境の設定が使用者に過度の負担となることなく、かつ、精度良く話者を特定することのできる話者照合装置及び方法を提供する。

【解決手段】 個人IDを入力し、個人IDに対応する登録話者の音声テンプレートを選択し、話者の音声を入力し、音声を分析して特徴パラメータを抽出し、入力された音声と登録話者の音声との照合距離、及び入力された音声と他の登録話者の音声との話者間距離を計算し、入力された音声が入力された個人IDに対応する話者と一致するか判定する話者照合装置及び方法であって、詐称者受理誤り率をしきい値として入力し、他の登録話者の音声と入力された音声との話者間距離に基づいて話者間距離の確率分布を求め、話者間距離の確率分布における詐称者受理誤り率の棄却域に、登録話者の音声と入力された音声との照合距離が含まれる場合に本人の音声であると判断する。

【選択図】 図3

出 願 人 履 歴 情 報

識別番号 [000005223]

1. 変更年月日	1996年 3月26日
[変更理由]	住所変更
住 所	神奈川県川崎市中原区上小田中4丁目1番1号
氏 名	富士通株式会社